

实时接收来自不同数据源的动态大尺度数据，高效并行地生成具有层次精度结构的**科学可视化数据**，便于高效查询。构造数据索引，能进一步提高数据查询速度。索引后的数据将被均匀分布存储于不同节点上，利于并行系统的负载平衡，提高硬件利用率。对数据作适当的冗余存储，确保重要数据不丢失。

利用 CPU 和 GPU 异构并行计算架构，进行科学可视化数据的高效绘制。对于二维数据，利用 GPU 硬件加速，以三角网格光栅化的方式绘制几何体，并将数据以贴图的方式附加到几何体上。对于三维体数据，利用可编程渲染管线硬件加速光线投射算法，进行体绘制。绘制结果将由 2-3 Swap 算法合成，然后压缩，发送到浏览器或者客户端软件。其中 CPU 负责完成数据和任务调度，以及高效压缩和数据传输的工作。

Talk on MPI with Yuxin Ma

Yuxin had done some research on MPI. And we had a brief talk on it.

Fonts in blue are the questions, and in green are answers. Some new questions occur in red

- 简介
 - 产生背景。
 - 开发年限。
 - 原理。
 - ◆ Message Passing Interface
 - 什么功能由他完成，什么不是他完成的。
- 许可证
 - 使用什么许可证？
 - 免费许可证吗？
 - 这个许可证要求开源吗？
 - 什么时候许可证是免费的？
- 状态
 - 社区活跃度。
 - 更新频率。
 - 健壮性。
- 安装
 - 从哪里下载？
 - 源代码大小。
 - 二进制包大小。
 - 如何进行绿色安装吗？
 - ◆ MPIs installing packages for Windows and Linux. It is very convenient. But there are some environment variables need be set. The installing packages can setup the environment variables automatically.

- 如何从源代码编译？

- ◆ To compile from the source code is very easy, just configure the path to install. The tools will do the rest work. And some features can be configured for advanced use.

- 编程

- 需要包含哪些库？

- ◆ STL
- ◆ POSIX
- ◆ Socket
- ◆ SSH

- Hello World 程序。

- ◆ <https://computing.llnl.gov/tutorials/mpi/>

```
#include "mpi.h"
#include <stdio.h>

int main(argc, argv)
int argc;
char *argv[]; {
    int numtasks, rank, rc;

    rc = MPI_Init(&argc, &argv);
    if (rc != MPI_SUCCESS) {
        printf ("Error starting MPI program. Terminating.\n");
        MPI_Abort(MPI_COMM_WORLD, rc);
    }

    MPI_Comm_size(MPI_COMM_WORLD, &numtasks);
    MPI_Comm_rank(MPI_COMM_WORLD, &rank);
    printf ("Number of tasks= %d My rank= %d\n", numtasks, rank);

    /***** do some work *****/

    MPI_Finalize();
}
```

- 教程资源多吗？

- 编译

- 需要给 IDE 安装插件吗？

- 需要使用它特有的预处理吗？

- 需要使用它特有的编译器吗？

- ◆ MPI has its own compiler MPICC.
- ◆ Will the compiler call native compiler?
- ◆ Will the compiler use native STL?

- ◆ Will the compile use its own STL?
- 运行
 - 需要给操作系统创建专用账号来运行程序吗？
 - ◆ MPI needs an account, which has access to remote node, to run programs on remote machine.
 - ◆ MPI works with SSH
 - 占用系统资源是多少？多少内存，多少硬盘空间，多少 CPU 时间。
 - 如何架设起整个集群？
 - 如何启动程序？
 - 可以动态添加节点吗？
 - ◆ ...unknown
 - 如何处理节点失败。
 - ◆ ...unknown
- 其他
 - 有节点监控功能吗？
 - ◆ ...unknown
- Tutorial
 - https://computing.llnl.gov/tutorials/parallel_comp/

Talk on Hadoop with Biao Zhu

Fonts in blue are the questions, and in green are answers. Some new questions occur in red

- 简介
 - 产生背景。
 - 开发年限。
 - 原理。
 - ◆ MapReduce
 - 什么功能由他完成，什么不是他完成的。
 - ◆ Including: distributed database, fault tolerance, cluster management...
 - ◆ Not included: load balance, fast result greeneration.
- 许可证
 - 使用什么许可证？
 - 免费许可证吗？
 - 这个许可证要求开源吗？
 - 什么时候许可证是免费的？
- 状态
 - 社区活跃度。
 - 更新频率。
 - 健壮性。
- 安装
 - 从哪里下载？
 - 源代码大小。
 - 二进制包大小。

- 如何进行绿色安装吗？
 - ◆ ...unknown
- 如何从源代码编译？
 - ◆ ...unknown
- 编程
 - 需要包含哪些库？
 - ◆ STL
 - ◆ SSH
 - Hello World 程序。
 - 教程资源多吗？
- 编译
 - 需要给 IDE 安装插件吗？
 - ◆ No
 - 需要使用它特有的预处理吗？
 - ◆ No
 - 需要使用它特有的编译器吗？
 - ◆ No
- 运行
 - 需要给操作系统创建专用账号来运行程序吗？
 - ◆ Hadoop needs an account, which has access to remote node, to run programs on remote machine.
 - ◆ Hadoop works with SSH
 - 占用系统资源是多少？多少内存，多少硬盘空间，多少 CPU 时间。
 - 如何架设起整个集群？
 - 如何启动程序？
 - 可以动态添加节点吗？
 - ◆ ...unknown
 - 如何处理节点失败。
 - ◆ Automatically
- 其他
 - 有节点监控功能吗？
 - ◆ Yes. A web based monitor.

Study BONIC

- Study Trac
- Download BONIC depends
- Download BONIC
- Read BOINC tutorial: <http://boinc.berkeley.edu/trac/wiki/ProjectMain>
 - Tutorial #1: Quick start
 - The BOINC server virtual machine
 - ◆ this article recommends setup the network with Bridged. But the network condition of our lab does not allow us to do like this. So I have to setup the network mode Host-only.

1. in VMware, open Virtual Machine Setting. and set the network mode to Host-only
2. in the virtual machine (Debian only), login as root.
 - a) nano /etc/network/interfaces
 - b) remove previous context and add:
allow-hotplug eth1
iface eth1 inet dhcp
 - c) save and exit
 - d) reboot
 - e) ifconfig eth1 down
 - f) ifconfig eth1 up

Now try to ping the virtual machine. It should be working (ping the host from the virtual machine does not work).

- Basic concepts
- Adapting applications
- Jobs and data
- Validation
- Security issues
- Project creation cookbook
- Give up. This framework is too complex to use.

Work delivery of server part for project climate

The server part has four processes:

- Data format convert. The data from simulation and satellite cannot be visualized directly, so we need to convert the data format. We need a person to do this job.
- LOD, index and storage. This is pre-process. Before the data is stored in database, we need to generate Level of Detail and index it. This process does not have time limitation, but it should be very reliable, stable, and the data is very large. So cloud computing fits this process best. Biao Zhu picked this job.
- Response. When a data request comes from client/browser, the parallel server should response. The request contains parameters, such as region, LOD level and data type. And the server will search the database, load and transfer the data to the client/browser. This process needs quick response. A request should be replied in several seconds. Besides, this process needs a kind of fault tolerance to ensure the request will be replied. But it will not do harm to our important data. So PVM fits this job best. I picked this job. Data processing, such as isosurface, is included in this process.
- Real-time rendering. When the server is ready for rendering, some rendering will be done on server, so as to support weak client/browser that does not has OpenGL ES or powerful GPU. In this process, fault tolerance is not necessary, because if one frame failed, continuing to render the next frame immediately will fix the problem. However, the rendering should be done very fast to meet interactive operation. This process will

be done on a small but powerful cluster, which is easy to upkeep. MPI fits this job best. And Yuxin Ma will pick this job.

- Spatial Database. We have two solutions:
 - We can choose to use MySQL or PostgreSQL. Both of them have Spatial Plugin. But Biao Zhu needs some time to get familiar with the database.
 - Biao Zhu will use HDFS to finish this job. But there will be more coding work for Biao Zhu.

We need to talk with somebody, who is familiar with the spatial database.